



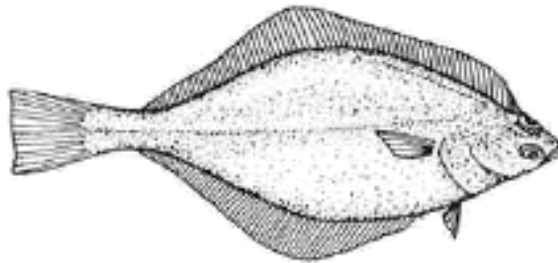
# NOAA Teacher at Sea Salish Sea DNA Project

## Lesson 07

## Constructing a Phylogenetic Tree

### Objectives

- Demonstrate the ability to perform UPGMA distance calculations
- Use UPGMA distance calculations to construct a phylogenetic tree of study populations
- Determine the phylogenetic relationships among the study populations



So far you have done quite a bit of work and the distance matrix you made in Lesson 06 will now be used to draw a phylogenetic tree. A phylogenetic tree is a reconstruction of evolutionary relationships among species or populations. The tree you will make is based on the genetic distance between our study populations. To draw your tree you will use a commonly used method. It is a distance matrix method (*good, since you just made a distance matrix*) called the unweighted pair-group method with arithmetic mean or UPGMA.

Making a tree is an iterative process of tree sketching, recalculating the genetic distance matrix, and more tree sketching. The example below will show you how to construct a tree using UPGMA. Read it carefully. When you have made your tree it will tell you the evolutionary relationship between populations. It does this because populations that share a branch point share a common ancestral population and are more closely related. The length of branches also contains information. Because microsatellites mutate randomly, as time passes, alleles are added and deleted from a population. If populations are reproductively isolated, each population will over time, accumulate a unique group of alleles and become more distant genetically. Therefore the length of the branch which is the genetic distance, is proportional to the evolutionary time. This means the length of branches between populations will illustrate the time since populations diverged. Logical, isn't it?

Read on to find out how to make an UPGMA tree!

### Making the Branches

To begin, you need a pairwise distance matrix and examine it to find the shortest distance. Using the distance matrix example from Lesson 06 (Figure 7.1) you see that the distance between populations **s** and **t** is the shortest.

**Figure 7.1**

English sole pairwise distance					
	p	q	r	s	t
q	0.818	X	X	X	X
r	0.563	0.363	X	X	X
s	0.635	0.713	0.765	X	X
t	0.645	0.522	0.622	0.195	X

Now you will draw the first branches of the tree. To draw the branch lengths first find the length of the branches. To do this divide the distance between **s** and **t** ( $d_{st}$ ) by 2 (branch length =  $d_{st}/2$ ). In this case  $0.195/2 = 0.098$ . Draw the branches like this:

**Figure 7.2**

t

Each horizontal line of this fine sketch is a branch and is 0.098 units long.

Populations **s** and **t** are now considered one population. Their distances from the others; **p**, **q**, and **r**, must be recalculated and the  $d_{st}$  value is dropped from the table because it has been used to make the first branches and all distances involving **s-t** are collapsed - added together and divided by 2. The boxed distances are the distances that will be combined and recalculated (Figure 7.3). Can you see why? Any distance not involving **s** and **t** are left as they are. So  $d_{pq}$ ,  $d_{pr}$ , and  $d_{qr}$  are the same as in the Figure 7.1.

**Figure 7.3**

English sole pairwise distance					
	p	q	r	s	t
q	0.818	X	X	X	X
r	0.563	0.363	X	X	X
s	0.635	0.713	0.765	X	X
t	0.645	0.522	0.622	0.195	X

The distance **st-p** is recalculated to 0.640, distance **st-q** to 0.618, and distance **st-r** is 0.694. After collapsing population **s** and **t** and recalculating the distances the distance matrix is as seen in figure 7.4

**Figure 7.4**

	p	q	r
q	0.818	X	X
r	0.563	0.363	X
s-t	0.640	0.618	0.694

Again examine the table to find the shortest distance. Looking at the table you see that the distance between populations **q** and **r** is the shortest at 0.363 (shaded cell Figure 7.5). These two populations will have branch lengths of  $0.363/2$  or 0.182. However, the tree is not resolved (See Figure 7.6).

**Figure 7.5**

	p	q	r
q	0.818	X	X
r	0.563	0.363	X
s-t	0.640	0.618	0.694

**Figure 7.6**

q

r

We know **s-t** and **q-r** are closely related but where does **p** fit in? Is it closer to **s-t** or **q-r**? And how do **s-t** and **q-r** connect to the tree overall? To find out we need to collapse **q-r** and recalculate the distance matrix. The boxed distances in Figure 7.4 will be collapsed resulting in the new matrix seen in Figure 7.6.

**Figure 7.6**

	p	q-r
q-r	0.690	X
s-t	0.640	0.656

So it turns out that **p** is closer to **s-t** with a branch length of  $0.640/2$  or  $0.320$ . Good, but now we need to collapse the last two distances into a single distance **stp – qr** to see how far **stp** is from **qr**. This distance is  $(0.690 + 0.656)/2 = 0.673$  with a branch length of  $0.673/2 = 0.336$ . I bet you want to draw the tree. Before we do, let's summarize the distances in a table (Figure 7.7). It will help us draw the tree.

**Figure 7.7**

Population Pair	Branch Length
s-t	0.098
q-r	0.182
st-p	0.320
stp-qr	0.336

With all the distances calculated we can now draw the tree. In this free-hand sketch the horizontal lengths of the branches are roughly the same as in the table.

You can see which populations share a common ancestral population and therefore which are more closely related. These relationships help biologist form new questions like:

- What caused populations **s** and **t** to split?
- How much gene flow is occurring between the populations?
- Do any of the populations have unique adaptations?
- How does this information affect the planning of a Marine Protected Area if English sole were to be included?
- If we were preparing a harvesting plan for English sole, should all or some of the populations be harvested?

Questions about how much time has passed, or in other words, how much time does each branch represent, can be answered but those methods are beyond this lesson.



You assignment is described on the next page.

**Activity 7.1 – Making an UPGMA Phylogenetic Tree**

Use the use the pairwise distance matrix you calculated in Activity 6.2 to make an UPGMA phylogenetic tree. Include each pairwise distance matrix iteration and your final tree must be made on graph paper with each branch length correctly scaled and labeled.